RESEARCH ARTICLE OPEN ACCESS

A Framework for Curved Videotext Detection and Extraction

Sudir.P[a], M.Ravishankar[b]
[a]*Asst.professor,Dept of E&C,SJCIT,Chikkaballapur,India*
[b]*Principal,VVIT,Mysore,India*

Abstract
Proposed approach explores a new framework for curved video text detection and extraction. The algorithm first utilizes a Gaussian filter based Color Edge Enhancement followed by a Gray level Co-occurrence matrix feature extraction method for text detection. Secondly, a Connected Component filtering method is utilized to generate clear localization result and at last, a Round Scan method is performed to extract curved text and generate binary result for recognition by OCR. Experiments on various curved video data and Hua's horizontal video text dataset shows the effectiveness and robustness of the proposed method.
Keywords*:*Curve Video text;Gaussian Filter;Gray level Co-occurrence matrix;Round scan Method.

I. Introduction

Due to tremendous advancement in Video technology, copious amount of video information available calls for the need for Semantic Video analysis and management. Since Video text is important supplemental content information, some of the Video processing techniques focus on text data embedded in videos. Video text detection and extraction is a vital task for applications such as Video Indexing and Content based video retrieval. Videotexts are classified into two types, first is Superimposed Text which is added as a part of editing process and the second is Scene text which appears naturally in a Scene captured by a camera. Although many methods exist, text detection and extraction is still a challenging task because of unconstrained colors, sizes, and alignments of the characters. Text detection and extraction methods can be divided into three subgroups: Connected Component (CC), Edge based and Texture-based. CC-based methods [1][2] groups small components into successively larger ones until all regions are identified in the image.CC-methods fail when text in video frame has low contrast and Non-Uniform Color. Edge based approach [3][4][5] requires text to have a better contrast to the background in order to detect the edges. So these methods exhibit poor result in case of complex backgrounds. Texture based methods [6][7] utilize various local texture descriptors, such as Gabor Filter[8] and Wavelet Filters [9].Since almost all of these methods rely on horizontal or Vertical alignment of Video text they fail in the case of Multi Oriented Video Texts. Recently some efficient algorithms have proposed specifically for Multi Oriented Video Texts. Shivakumara et al.[10] proposed a method for detecting arbitrarily oriented text based on Morphological approach but the method fails in case of complex background. Yong et al.[11] proposed a corner and skeleton based method for arbitrarily oriented text which is quiet robust only when the corners are detected effectively. It fails in case of low resolution Videos. Shivakumara et al. [12] proposed an arbitrarily oriented scene text detection method which extracts text lines of any orientation based on Gradient Vector Flow (GVF) and Neighbor component grouping. However the proposed method does not give good accuracy for less spaced text lines. Shivakumara et al. [13] proposed a curved text detection scheme which uses enhancement criterion and quad tree approach. Although efficient in detection proposed method fails to eliminate false positives. Clearly all of the above methods though efficient in detection fail to produce an OCR ready input. Hence in this paper an efficient text detection and extraction scheme is proposed for Curved Videotexts. The main contributions of the proposed method are as follows: 1) A Gaussian Filter based Video frame Enhancement Model 2) A Gray level Co-occurrence matrix feature extraction method for text detection 3) A Round scan method for partial and fully circular videotext recognition is proposed which segments curve texts and produces binarized text segments for recognition by OCR.

II. Proposed Method

The proposed method consists of three steps: Filter based Color edge enhancement, Text detection by Extracting Standard deviation as feature from block wise Gray level Co-occurrence matrix generated, A Round scan method for circular and semicircular text extraction.

2.1 Gaussian Filter based Color Edge Enhancement

The purpose of use of text edge enhancement procedure is to highlight fine detail in an image or to restore, at least partially, text detail that has been blurred. Text Edge enhancement involves sharpening the text edges with respect to their background. In the proposed method first each set of R, G and B-bands are smoothed by a 3x3 Gaussian kernel. The reason for choosing Gaussian kernel is that it is an efficient filter for removing noise drawn from a normal distribution and Gaussian functions are rotationally symmetric in two dimensions hence it will not bias subsequent edge detection in any particular direction. The degree of smoothening is governed by variance s. A larger s implies a wider Gaussian filter and greater smoothening. The Gaussian filter in the continuous space is given b

$$h(m,n) = \left(\frac{1}{\sqrt{2\pi\sigma}} e^{\frac{m^2}{2\pi^2}}\right)\left(\frac{1}{\sqrt{2\pi\sigma}} e^{\frac{n^2}{2\pi^2}}\right) \quad (1)$$

Let I_x be original band where x ∈ (r, g, b) and J_x be the Gaussian filter smoothed band. Then the enhanced band E_x is given by the formula.

$$E_x=(1+A)I_x-AJ_x \quad (2)$$

Where A is an Amplification factor (A=1 and σ=5 in our experiment). Let F be the Edge Enhanced Video frame obtained by concatenating E_r ,E_g and E_b. The result obtained after the enhancement procedure is shown in Fig.1(b).

2.2 Text detection method

Proposed approach which is texture-based is based on the a statistical method of examining texture that considers the spatial relationship of pixels which is the gray-level co-occurrence matrix (GLCM), also known as the gray-level spatial dependence matrix. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix.
Given an image I, of size M×M, the co-occurrence, matrix G can be defined as

$$G(i,j) = \sum_{x=1}^{M}\sum_{y=1}^{M}\begin{cases}1 & I(x+\Delta x, y+\Delta y) = j\\ 0 & otherwise\end{cases} \quad 3)$$

Here, the offset (Δx, Δy) specifies the distance between the pixel-of-interest and its neighbour. The offset (Δx, Δy) parameterization makes the co-occurrence matrix sensitive to rotation. Choosing an offset vector, such that the rotation of the image is not equal to 180 degrees, will result in a different co-occurrence matrix for the rotated image. Standard deviation feature is extracted from 4 gray co-occurrence matrix at 4 different rotations given by [0 Δ] for 0◦: G horizontal, [-Δ, Δ] for 45◦: G right diagonal, [-Δ 0] for 90◦: G vertical, and [-Δ − Δ] for 135◦: G left diagonal.
The input video frame is partitioned into n^2 number of blocks. For each block B,let B(i,j) be the $(i,j)^{th}$ entry in a normalized GLCM. The mean and standard deviations for the rows and columns of the matrix are

$$\mu_x = \sum_i \sum_j i.B(i,j),\ \mu_y = \sum_i \sum_j j.B(i,j) \quad (4)$$

$$\sigma_x = \sum_i \sum_j (i-\mu_x)^2 B(i,j)\ ,\ \sigma_y = \sum_i \sum_j \left(j-\mu_y\right)^2 B(i,j) \quad (5)$$

Standard deviation for every block for all four rotations are calculated and the maximum value among all these are considered. The standard deviation feature value are threshold by Otsu's method [14], which chooses the threshold to minimize the interclass variance of the black and white pixels. The result is a black and white image where white color signifies probable text regions. The block in the original image F is retained as it is if its corresponding threshold image value is 1 else the block is suppressed by assigning 0 to all the pixels in the block. The resultant image is shown in Fig.1(c).

$$f(x,y)=\begin{cases} f(x,y) & \text{if } B(i,j)=1\\ 0 & \text{otherwise}\end{cases} \quad (6)$$

2.3 False Positive removal method

False Positive elimination is executed based on size of CC and height of CC since it is a known fact that all the characters in a circular text will be of almost of same height. Height of a CC of circular text is chosen as maximum among length of CC and width of CC. In our experiment we ignore CCs whose size is less than 200 and height less than 20 and height greater than 70.The result of false positive elimination is shown in Fig.1 (e, k, l, m, n, o).

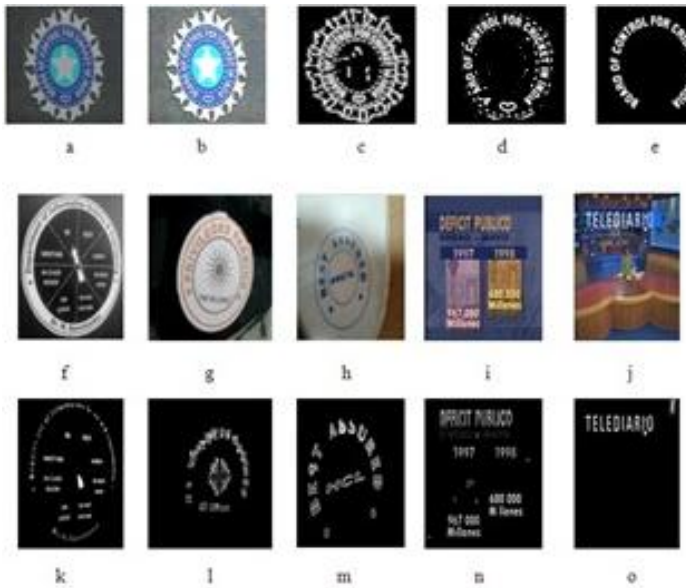


Fig.1. Curved Text Detection and Localization result (a,f,g,h,i,j)-Video Frame b. Enhanced Frame (c) Text Detection result without Enhancement (d) text detection result with enhancement (e,k,l,m,n,o)- Result after False positive elimination

2.4 Text Extraction and recognition procedure

The proposed procedure aims to extract the curve text for recognition using OCR from the result obtained after false positive elimination. Two types of curve texts have to be dealt with for segmentation; the first case is where the curve text covers more than 75% of the frame which we call as fully circular text (Fig.2e) and curve text which covers less than 50% of the frame which we call as partial circular text (Fig.2g). In order to segment in both cases one needs to find a origin point of scan. The procedure to find origin point of scan is as follows:

- *Fully Circular text*: Obtain Horizontal projection as shown in Fig.2c.Find first non zero X-axes value(X1) by scanning X-axes value from left to right and first nonzero X-axes value(X2) by scanning X-axes value from right to left. Similarly obtain Vertical projection and find first nonzero Y-axes value (Y1) by scanning Y-axes value from top left to bottom and first non zero Y-axes value (Y2) by scanning Y-axes value from bottom to right as shown in Fig.2b. Origin point of scan C(x, y) is found by using the formula and is shown in Fig.2e.

$$x = \frac{X2 - X1}{2} \quad \text{and} \quad y = \frac{Y2 - Y1}{2} \qquad (7)$$

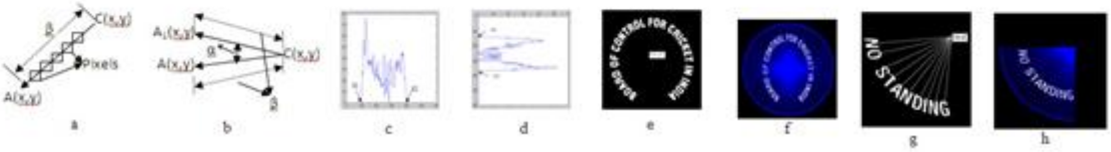
- *Partially Circular text*: Choose a Coordinate manually as Origin point of scan C(x, y) such that it is at equidistant from the centroid of all CCs as shown in Fig.2h.

Round scan method: In order to extract curved text first the Origin point of scan is chosen and an extreme point A (i, j) is found at a distance β chosen heuristically so as to cover entire text region. All the pixel points on the linear path from $C(x_1, y_1)$ to $A(x_i, y_i)$ is extracted and stored as shown in Fig.2a. The next point A1(x, y) is found as per equation 10 and again all pixel values are extracted and stored.

$$x=x_1-\beta\sin(\alpha\pi/180),\ \ y=y_1-\beta\sin(\alpha\pi/180), \qquad (8)$$

Above procedure is repeated for full 360^0 as shown in fig.2f by finding new extreme points and all the pixels in the path from origin point of scan to extreme points are extracted and stored and the resultant image obtained are shown in Fig.2.

Fig.2. (a) Extract each scan line pixel values (b) to find new scan line (c) Horizontal Projection (d)Vertical



Projection (e) Origin point of scan (f) Scan lines (g) Origin point of scan (h) Scan lines

Recognition of curved text in Video frames is a challenging task due to complex background and arbitrary orientations. There are few methods [16, 17, 18] which extract curved text from natural scene images. However, the scope of these methods is limited to image text detection and extraction but not binarization and recognition.

Shivakumara et al [18] present a novel method of multi oriented text recognition in scene images using HMM and performs segmentation free recognition in curved text image. Proposed method is the first of its kind which linearizes and binarizes the curve Video text and provides input to OCR for recognition. The extracted text using Round scan method is binarized using Otsu Threshold and Recognized using Tesseract OCR [19]. Sample results are shown in Fig.3.

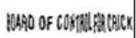
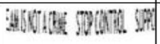
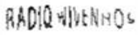
Video frame	Extracted text	OCR output	Video frame	Extracted text	OCR output
		IMO OF CONORMCRICKET INPI			EN5MVAUN€ STOI'CONIR01. SVCI
		RAOIQ VIVLNRO			VO STMOINC

Fig.3. Sample Curved Text Extraction and Recognition results

III. Results

The proposed approach has been tested on 20 mpeg-1 video frames with 320x240 pixel resolution as there is no benchmarked dataset on Curved video text data, available in the literature. It contains 92 words and 670 characters. We also use a small dataset of 45 video frames, which is available, publicly [10] to evaluate the performance of the method. We choose four methods for comparison of which first is a recently developed method by Shivakumara et al.[13] which address curved text detection in video through an enhancement criterion and use of quad tree. Second is Laplacian method [10] which works well for non-horizontal Video text by combining Laplacian and Fourier. Third is Bayesian method [21] which explores Bayesian classifier and gradient. Fourth is a method for arbitrary text detection by Sharma et al. [16] which proposes gradient directions and two stage grouping. Following performance criterion has been defined for horizontal, vertical and multioriented videotext (except curved text).

- Truly Detected Block (TDB): A detected block that contains at least one true character. Thus, a TDB may or may not fully enclose a text line.
- Falsely Detected Block (FDB): A detected block that does not contain text.
- Text Block with Missing Data (MDB): A detected block that misses more than 20% of the characters of a text line (MDB is a subset of TDB).

The performance measures are defined as follows.

$$1)\ Recall(R) = \frac{TDB}{ATB} \qquad 2)\ Precision(P) = \frac{TDB}{TDB + FDB} \qquad 3)\ F-measure(F) = \frac{2PR}{P+R}$$

(9)

Average Processing Time (APT) per frame is also considered as a measure to evaluate the proposed method.

1.1. Performances on Hua's data

A sample frame from Hua's data is shown in Fig.1(i,j) and the text detection results of the proposed method for the frame in Fig.1(n,o).The proposed method detected almost all text properly. The Comparison results of the proposed method and existing method are reported in Table 1. The results show that our method is robust to the orientation, perspective, color, and lighting of the text object. It detected most text objects successfully.

1.2. Performance Analysis on Curved data

Sample results of the proposed method for curved text extraction versus the Quad tree based method is shown in Fig.4 where (a,d,g) shows input frames having different background and fonts and (b,e,h) shows the results given by the proposed method and (c,f,i) shows the result for the Quad tree based method. Fig.4 shows that the proposed method extracts curved text successfully. Table 1 and 2 show that the proposed method outperforms the existing methods in terms of recall, precision-measure and misdetection rate. The main reason for poor results of the existing methods is that the existing methods are developed for horizontal and non-horizontal text detection but not for curved video text detection. Table 3 shows the recognition accuracy obtained on the curved dataset. The Accuracy is fairly good and there is still scope for improvement in this regard.

Fig.4. Text Detection result (a,d,g) Input Frames,(b,e,h) Proposed method result,(c,f,i) Quad tree based method

Table 1. Comparison of proposed method with Existing Detection Methods for Hua's database

Methods	R	P	F	MDR	APT(secs)
Proposed	0.87	0.9	0.9	0.06	1.7
Quad tree	0.82	0.88	0.84	0.06	2.0
Sharma	0.88	0.77	0.82	0.32	9.0
Bayesian	0.87	0.85	0.85	0.18	5.6
Laplacian	0.93	0.81	0.87	0.07	11.7

Table 2. Performance of the Proposed and Existing Detection Methods on Curved Data

Methods	R	P	F	MDR	APT(secs)
Proposed	0.82	0.87	0.84	0.20	1.7
Quad tree	0.80	0.83	0.81	0.25	2.3
Sharma	0.73	0.88	0.79	0.28	10.3
Bayesian	0.59	0.52	0.55	0.27	12.1
Laplacian	0.55	0.68	0.60	0.42	9.9

Table 3. Character and Word Recognition Performance by Proposed Recognition Method

Dataset	Accuracy	
	Character	Word
Own Curved Data	49	43

IV. Conclusion

In this paper a novel method is proposed for curved video text detection and extraction which uses the standard Deviation feature embodied in Gray Co-occurrence matrix obtained from Gaussian filter enhanced Video frame. A Novel Round Scan method is performed to extract curved text and generate binary result for recognition by OCR. Some experimental results and performance comparisons with standard methods are reported in detail, thereby confirming that our method is capable of robustly handling Curved video text. Our future study will mostly be focusing increasing Curved video text recognition rate.

References

[1] A.K. Jain and B. Yu. (1998) Automatic Text Location in Images and Video Frames, *Pattern Recognition,*Vol.31 (12), pp. 2055-2076.

[2] Lienhart,R., Stuber,F.Automatic.(1996).Text Recognition in Digital Videos, *Proceedings of SPIE Image and VideoProcessing* IV 2666, pp. 180-188.

[3] Agnihotri,L., Dimitrova,N.(1999)Text Detection for Video Analysis*, IEEE Workshop on CBAIVL*, pp. 109113.

[4] Agnihotri,L., Dimitrova,N., Soletic,(2002),M.Multi-layered Videotext Extraction Method, *IEEE International Conference on Multimedia and Expo (ICME)*, Lausanne (Switzerland), August 26-29, 2002.

[5] Hua,X.-S., Chen, X.-R. et al.(2001).Automatic Location of Text in Video Frames, *Intl Workshop on Multimedia Information Retrieval* In conjunction with ACM Multimedia.

[6] Kim, H-K. et al.(1996).Efficient Automatic Text Location Method and Content-based Indexing and Structuring of Video Database, *Journal of Visual Communication and Image Representation*, vol.7,No.4, pp.336-344.

[7] Li,H., Doermann,D, Kia,O.(2000).Automatic Text Detection and Tracking in Digital Video, *IEEE Trans. on Image processing*.vol.9, No.1, pp. 147-155.

[8] Jain,A.K., Bhatarcharjee,S.(1992).Text Segmentation using Gabor Filters for automatic document processing, *Machine Vision and Application*, vol.5 pp. 169 -184.

[9] Li,H., Doermann,D, Kia,O.(1998).Automatic Text Detection and Tracking in Digital Video, Univ. of Maryland, College Park, Tech.Reps. LAMP-TR-028, CAR-TR-900.

[10] P. Shivakumara, T. Q. Phan and C. L. Tan.(2011).A Laplacian Approach to Multi-Oriented Text Detection in Video, *IEEE Trans. on PAMI*, pp. 412-419.

[11] Yong Zhang, Jianhuang Lai.(2012) Arbitrarily Oriented Text Detection Using Geodesic Distances Between Corners and Skeletons, *1st International Conference on Pattern Recognition (ICPR),*Tsukuba, Japan.

[12] Palaiahnakote Shivakumara, Trung Quy Phan, Shijian Lu and Chew Lim Tan.(2013).Gradient Vector Flow and Grouping based Method for Arbitrarily Oriented Scene text Detection in Video Images, *Circuits and Systems for Video Technology, IEEE Transactions on* ,Volume:23 , Issue: 10.

[13] P.Shivakumara, H.T.Basavaraju, D.S.Guru, and C. L.Tan.(2013).Detection of Curved Text in Video: Quad Tree based Method, *ICDAR.*

[14] N. Otsu.A.(1979). Threshold Selection Method from GrayLevel Histograms, *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. 9,, pp. 62-66.

[15] H.K. Koh, W. Shen, B. Shuter and A.A. Kassim.(2009).Segmentation of kidney cortex in MRI studies: a constrained morphological 3D h-maxima transform approach,*International Journal of Medical Engineering and Informatics,* Vol. 1, No. 3, pp. 330-341.

[16] N. Sharma, P. Shivakumara, U. Pal, M. Bluemenstein and C. L. Tan.(2012).A New Method for Arbitrarily-Oriented Text Detection in Video, In Proc. *DAS*, pp. 74-78.

[17] C. Yao, X. Bai, W. Liu, Y. Ma and Z. Tu.(2012).Detecting Texts of Arbitrary Orientations in Natural Images,*CVPR.*

[18] Sangheeta Roy, Partha Pratim Roy, Palaiahnakote Shivakumara, Umapada Pal ,Chew Lim Tan, (ACPR2013).HMM based Multi Oriented Text Recognition in Natural Scene Image, *The 2nd Asian Conference on Pattern Recognition*.

[19] http://code.google.com/p/tesseract-ocr/

[20] X. S Hua, L. Wenyin and H.J. Zhang.(2004).An Automatic Performance Evaluation Protocol for Video Text Detection Algorithms*, IEEE Trans. on CSVT*, 498-507.

[21] P. Shivakumara, R. P. Sreedhar, T. Q. Phan, S. Lu and C. L. Tan.(2012).Multi-Oriented Video Scene Text Detection through

Bayesian Classification and Boundary Growing,*IEEE Trans. on CSVT,*1227-1235.